

ASSIGNMENT METHODS IN THE THREE-FORM PLANNED MISSING DATA DESIGN

By

Terrence D. Jorgensen

Submitted to the graduate degree program in Psychology
and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Arts.

Chairperson: Todd D. Little, Ph.D.

Wei Wu, Ph.D.

Paul E. Johnson, Ph.D.

Date Defended: May 29th, 2013

The Thesis Committee for Terrence D. Jorgensen
certifies that this is the approved version of the following thesis:

ASSIGNMENT METHODS IN THE THREE-FORM PLANNED MISSING DATA DESIGN

Chairperson: Todd D. Little, Ph.D.

Date approved: May 29th, 2013

Abstract

The 3-form planned missing data design allows researchers to measure more items from more participants on more occasions using the same budget as a complete data design. It also reduces burden, fatigue, and response reactivity. After randomly assigning participants to complete 1 of 3 forms on the first occasion, for subsequent measurements researchers might assign participants to the same or a different form, or use random assignment again. I discuss potential advantages and drawbacks of each approach, including a simulation to compare bias across methods. Results indicate negligible differences between assignment methods only in the absence of retest effects. Reduction of bias due to retest effects makes it preferable to systematically assign different forms over time.

Acknowledgements

I would like to thank each of my committee members, Drs. Todd Little, Wei Wu, and Paul Johnson, for their expertise, time, guidance, and support. I am grateful for the statistical resources and technological support provided by the University of Kansas's Center for Research Methods and Data Analysis. I would also like to thank all members of the missing data workgroup, where the seeds of this project were sown. This research was funded in part by a grant (NSF-1053160) from the National Science Foundation (Wu & Little, co-PIs), and in part by the Center for Research Methods and Data Analysis (Todd Little, Director), University of Kansas.

Table of Contents

PART I: Background and Motivation.....	1
Rationale for Planned Missing Designs.....	2
The Three-Form Design.....	4
Retest Effects.....	8
PART II: Equivalence of Assignment Methods in the Absence of Practice Effects.....	8
Monte Carlo Simulation Design.....	8
Results	12
Parameter Estimates.....	13
Improper Solutions	19
PART III: Differences among Assignment Methods in the Presence of Practice Effects.....	19
Modeling Practice Effects.....	19
Monte Carlo Simulation Design.....	21
Results	23
PART IV: Conclusions.....	24
Suggestions for Applied Researchers.....	24
Limitations and Future Directions.....	28
References.....	31
Appendix A: R Syntax for One Condition in Part II.....	34
Appendix B: R Syntax for One Condition in Part III.....	37

List of Tables and Figures

Table 1: Schematic for a Three-Form Planned Missing Design.....	4
Table 2: Range of Bias, MSE, and 95% Coverage in Six Planned-Missing Conditions.....	15
Figure 1: Missing data pattern associated with assigning the same form over time.....	6
Figure 2: Missing data pattern associated with assigning different forms over time.....	6
Figure 3: Missing data pattern associated with assigning forms randomly over time.....	7
Figure 4: Data-generating and analysis model for Part II.....	10
Figure 5: Items within a construct are evenly distributed across blocks.....	11
Figure 6: Variability in estimated factor loading across conditions.....	16
Figure 7: Variability in estimated factor regression across conditions.....	17
Figure 8: Power to detect latent regressions increases with effect size.....	18
Figure 9: Assignment of items to blocks at one occasion.....	22
Figure 10: Bias in latent means and intercepts due to practice effects over time.....	25

Assignment Methods in the Three-Form Planned Missing Data Design

PART I: Background and Motivation

In social and behavioral science research, missing data are practically unavoidable but need no longer be an insurmountable problem. Full information maximum likelihood (FIML; Anderson, 1957) estimation and multiple imputation (MI; Rubin, 1987) are thoroughly vetted methods that use all observed information to obtain unbiased point and *SE* estimates when the missingness mechanism is ignorable (Baraldi & Enders, 2010). Schafer and Graham (2002; also Enders, 2010, Graham, 2012; van Buuren, 2012) provide a thorough discussion of why these methods are preferable to listwise or pairwise deletion, mean substitution, or regression-based single imputations, as well as an overview of software packages in which these state-of-the-art methods are available. With the computing power now available to implement such methods, missing data need not be avoided at all cost. In fact, missing data can be incorporated into study designs making them more cost effective, giving researchers a way to exercise a great deal of control over the proportion and, more importantly, the mechanism of missingness in their data. Graham, Hofer, and MacKinnon (1996) advocated for a three-form design as a way to make an efficient compromise between measuring as many relevant variables as necessary and preserving the quality of the data by reducing burden and preventing fatigue (see also Graham, Taylor, Olchowski, & Cumsille, 2006). Harel, Stratton, and Aseltine (2011) also showed that planned missingness reduces assessment reactivity and limits the amount of unplanned missing data.

My primary goal for this thesis is not to introduce the three-form planned missing data design, but rather to provide guidance on how to apply a three-form planned missing data design in a longitudinal setting. In so doing, I focus on key considerations not yet addressed in the existing literature. I begin with a brief discussion of missingness mechanisms to provide a

justification for using planned missing designs in general, followed by a discussion of such designs, with special emphasis on the three-form design in a developmental context. The primary goal of this article is to illustrate how different methods of assigning forms to participants across measurement occasions can affect the efficiency of parameter estimates. I will present results from two Monte Carlo simulations designed to investigate the consequences of different assignment methods. The first, in which practice effects are absent, demonstrates the equivalence of different assignment methods. In contrast, the second study highlights differences among assignment methods in the presence of practice effects. I conclude with recommendations for applied researchers and for future methodological research.

Rationale for Planned Missing Designs

The ability of FIML or MI to provide unbiased point and *SE* estimates in the presence of missing data hinges on the mechanism of missingness (Enders, 2010; Graham et al., 2006). In order for missingness to be ignorable, observations must be missing at random, either completely or after taking into account any variables that are related to missingness. Data are missing completely at random (MCAR) when the reason why the observations are missing is completely unrelated to either the observed or missing data, such as when data are lost due to random computer error. Unplanned missing data are likely to be missing for some systematic reason, but they can still be missing at random (MAR) when other measured variables are either the cause of or related to the cause of missingness. If such related variables are not measured, or if data on a variable are missing due to values of the variable itself—such as when people with particularly low or high income fail to report their income—then data are missing not at random (MNAR).

Enders (2010) details how all available information is used to estimate relationships in the presence of data that are missing due to MAR or MCAR mechanisms. Data are inevitably

missing in the course of collection, and though there are tests to distinguish MAR from MCAR missingness (e.g., Little, 1988), there is no way to be certain about the missingness mechanism. To raise the likelihood of MAR data requires anticipating which data might end up missing and diligently planning to measure variables most likely to be related to the pattern of missingness. These precautions are prudent even when using a planned missing design. With planned missing data, researchers can completely control the mechanism of those missing data. By randomly assigning participants to conditions in which they do not respond to certain items—under the same rationale for random assignment to experimental conditions—the mechanism is by definition MCAR, which does not require the measurement of related variables.

Although controlling missingness does not prevent additional data from systematic missingness, it can decrease the amount of missingness related to unplanned mechanisms (Harel et al., 2011). Using planned missing data designs, such as the three-form design (explained in the next section), brings with it numerous advantages. In cross-sectional designs, response burden and fatigue are reduced because the protocol that each person completes is a shortened form. Costs are also reduced because the time commitment per participant (and per research assistant) is reduced. Particularly for large-scale studies, planned missingness designs result in more cost-effective and efficient data collection (Harel et al., 2011). One such design is a multiform design, which assigns the fractions of an entire battery of measurements to different subsets, only some of which are assigned to different forms, so that each participant will complete a subset of all measurements (more detail is provided in the next section). In longitudinal designs with a multiform component, assessment reactivity—change in participants' behavior in response to the assessment protocol—is reduced (Harel et al., 2011). Assessment reactivity is but one form of retest effect, but any kind of retest effect can be minimized with planned missing protocols.

Optimal guidance on how to best implement a multiform design is an area of ongoing research. The primary goal of this thesis is to provide such direction for implementing multiform designs, specifically in longitudinal research.

The Three-Form Design

The three-form design is the simplest type of multiform design. In cross-sectional studies it is essentially a way to sample variables so that a large number of variables can be measured within a limited time frame or attention span. The design divides a large number of items (e.g., survey questions) into four subsets, a common block labeled X and three partial blocks labeled A, B, and C (see Table 1). The block X is given to all participants, and it should contain essential variables (e.g., demographics, assignment variables). Two of the A, B, and C blocks are randomly selected and assigned to each participant. As a result, each participant is administered one of the three combinations (XAB, XBC, and XAC) and has missing data for items in the fourth block. For example, if each block contains the same number of variables and each of the combinations is administered to one third of the participants, then 25% more items can be administered than complete data design given the same time constraints.

Table 1

Schematic for a Three-Form Planned Missing Design.

Form	Common Set X	Variable Set A	Variable Set B	Variable Set C
1	25% of items	25% of items	25% of items	Missing
2	25% of items	25% of items	Missing	25% of items
3	25% of items	Missing	25% of items	25% of items

Note. Proportions of variables in each block need not exactly match this schematic.

This three-form design has many advantages. As mentioned above, it provides a dataset

with more variables (making short-form versions of scales less necessary), which allows more research questions or represents constructs more completely. Alternately, if the number of items is not increased, this design can reduce the effect of fatigue in a data collection procedure and improve the data quality as only a subset of variables is given to each participant. Finally, by imposing planned MCAR missingness, unplanned missingness is reduced, increasing the validity of the data.

The three-form design can also be used to sample variables across measurement waves in longitudinal studies. When the three-form design is used to sample variables at each wave, a decision needs to be made regarding how to assign the three forms to participants across waves. For example, should participants who were given XAB at the first measurement occasion also be given the same form (XAB) at all of the subsequent measurement occasions, or should steps be taken to assign different forms (either XBC or XAC) on subsequent occasions, either systemically or by random assignment? The assignment strategy that provides the best combination of the following factors would be deemed optimal: accuracy of the point and *SE* estimates for key parameters, power to detect the key parameters, and cost effectiveness.

The optimal strategy is not immediately obvious and may vary depending on the research question, what model is specified, and which parameters are of primary interest. I do not hypothesize that the point estimates of parameters to depend on assignment method, but because the different assignment schemes yield different patterns of missingness, one might expect them to yield different relative efficiency for estimating those parameters. For instance, assigning the same form over time (see Figure 1) might favor the precision (i.e., smaller *SE*) of estimates of covariances between the same variable(s) across waves, whereas assigning different forms (see Figure 2) would favor the precision of estimates of covariances between different variables

across waves. Following these expectations, if the goal were to establish longitudinal measurement invariance in a CFA model, for example, then assigning the same form over time would yield the most precise estimates of key parameters in the measurement model. A structural regression model of longitudinal mediation would, however, point to different forms as the preferred method of assignment.

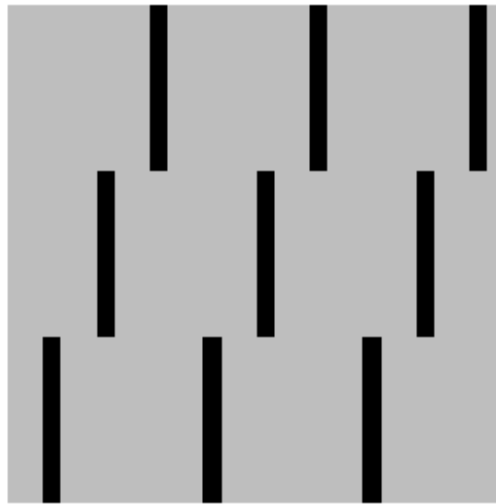


Figure 1: Missing data pattern associated with assigning the same form over time.

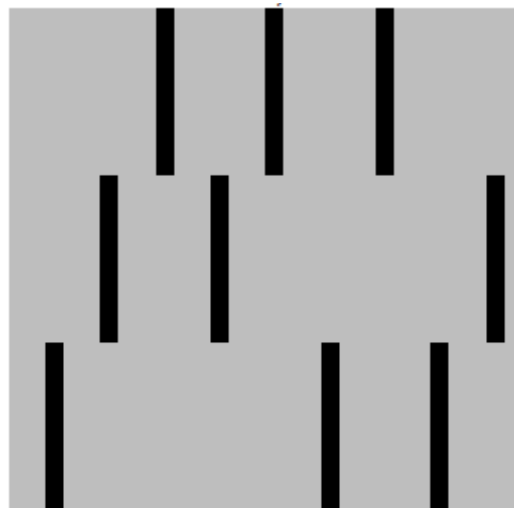


Figure 2: Missing data pattern associated with assigning different forms over time.

Assigning forms randomly on each occasion would yield mixtures of participants seeing

the same form on consecutive occasions and participants seeing different forms (see Figure 3), but it does not necessarily follow that it would yield “the best of both worlds” as far as relative efficiency of estimation is concerned. As the number of measurement occasions (T) increase, the number of possible missing data patterns (M) increases, such that $M = 3^T$. Using a three-forms design would therefore yield 27 missing data patterns in a three-occasion design, 81 missing data patterns in a four-occasion design, 243 missing data patterns in a five-occasion design, 729 missing data patterns in a six-occasion design, etc.

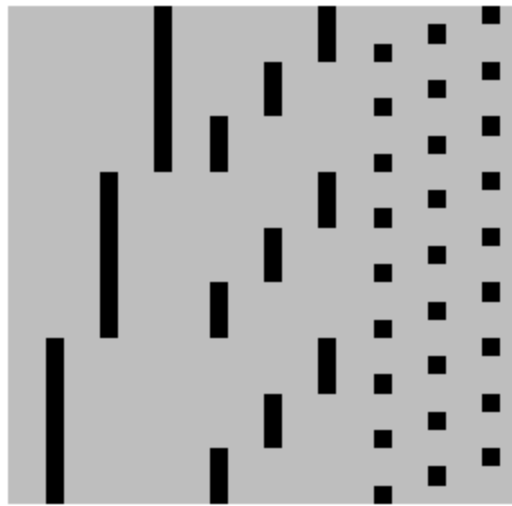


Figure 3: Missing data pattern associated with assigning forms randomly over time.

These are merely lower limits for the total number of missing data patterns because unplanned missing data will yield additional patterns within any planned missing data pattern. Thus, randomly assigning forms on each occasion might yield at most one person per missing data pattern when, for example, a sample of $N = 200$ is measured on four occasions. Although this might not necessarily lead to convergence problems, calculating the ML discrepancy function (using FIML) for each missing data pattern would necessarily lead to longer times until convergence as the number of missing data patterns increases. Without an explicit advantage to randomly assign forms on each occasion instead of systematically assigning different forms over

time, the longer wait time until convergence would make the random-assignment method less preferable than the systematic method, which yields as few missing data patterns as would assigning the same form over time (compare Figures 1 and 2).

Retest Effects

As mentioned earlier, retest effects are justifiably expected when making repeated measurements. Assigning different forms over time would decrease the number of variables that participants see on consecutive occasions, thus decreasing the degree to which retest effects manifest in variables not included in the X block. I therefore hypothesize that assigning different forms over time would be the preferred method in the presence of any kind of retest effects, and that randomly assigning forms at each wave would not be as effective at suppressing retest effects as would a systematic method to ensure that participants would not see the same form twice, at least not until they had seen each other form first.

I aim in my thesis to provide some insights into the issue of assignment methods by comparing the different assignment strategies in the context of a commonly used type of longitudinal model: autoregressive latent-variable panel models. The first study (in Part II) is an investigation of the effect of assignment method in the absence of retest effects, which is useful to establish a sort of baseline of differences between assignment methods. In the second study (in Part III), I introduced retest effects to investigate differences between methods in a more realistic situation.

PART II: Equivalence of Assignment Methods in the Absence of Practice Effects

Monte Carlo Simulation Design

In my first Monte Carlo simulation study, I used a three-wave, three-factor autoregressive cross-lagged longitudinal model (Figure 4), which can be used for mediation (Cole & Maxwell,

2003). Each latent variable had three indicators at each wave. In the population and analysis models, all autoregressive and cross-lagged paths were of order one (e.g., autoregressive paths of X_2 on X_1 , but not X_3 on X_1) and cross-lagged paths only went from X to M and from M to Y , representing full mediation. The population measurement model (not depicted in Figure 1) included Lag-1 residual correlations of 0.2 across time (e.g., the residual covariance from the first indicator of X at Time 1 and the first indicator of X at Time 2) and a Lag-2 residual correlation of 0.04. The analysis model estimated residual covariances at Lags 1 and 2. All indicator intercepts and latent means were zero in the population, and latent residual variances were fixed such that all total latent variances were equal to one. Sample size in all conditions was fixed to $N = 540$.

The following model parameters were manipulated: factor loadings, within-time latent correlations (or residual covariances after Time 1), autoregressive paths, and cross-lagged paths—each of which was equal across time in the population model (e.g., all factor loadings at each time had the same population value). Strong factorial invariance was established by constraining factor loadings and intercepts to equality across time in the analysis model. I identified the analysis model by fixing latent variances at Time 1 to one and latent means at Time 1 to zero. Factor loadings varied from 0.7 to 0.85, by increments of 0.05; within-time correlations varied from 0.2 to 0.5 by increments of 0.1; autoregressive paths varied from 0.4 to 0.9 by increments of 0.1; and cross-lagged paths varied from 0.0 to 0.4 by increments of 0.1. This resulted in $4 \times 4 \times 6 \times 5 = 480$ population models; however, because residual variances were constrained to make each total variance equal to one, some combinations of conditions resulted in impossible population values given that the total variances were one (e.g., a latent residual variance < 0). These conditions were omitted, resulting in 436 population models.

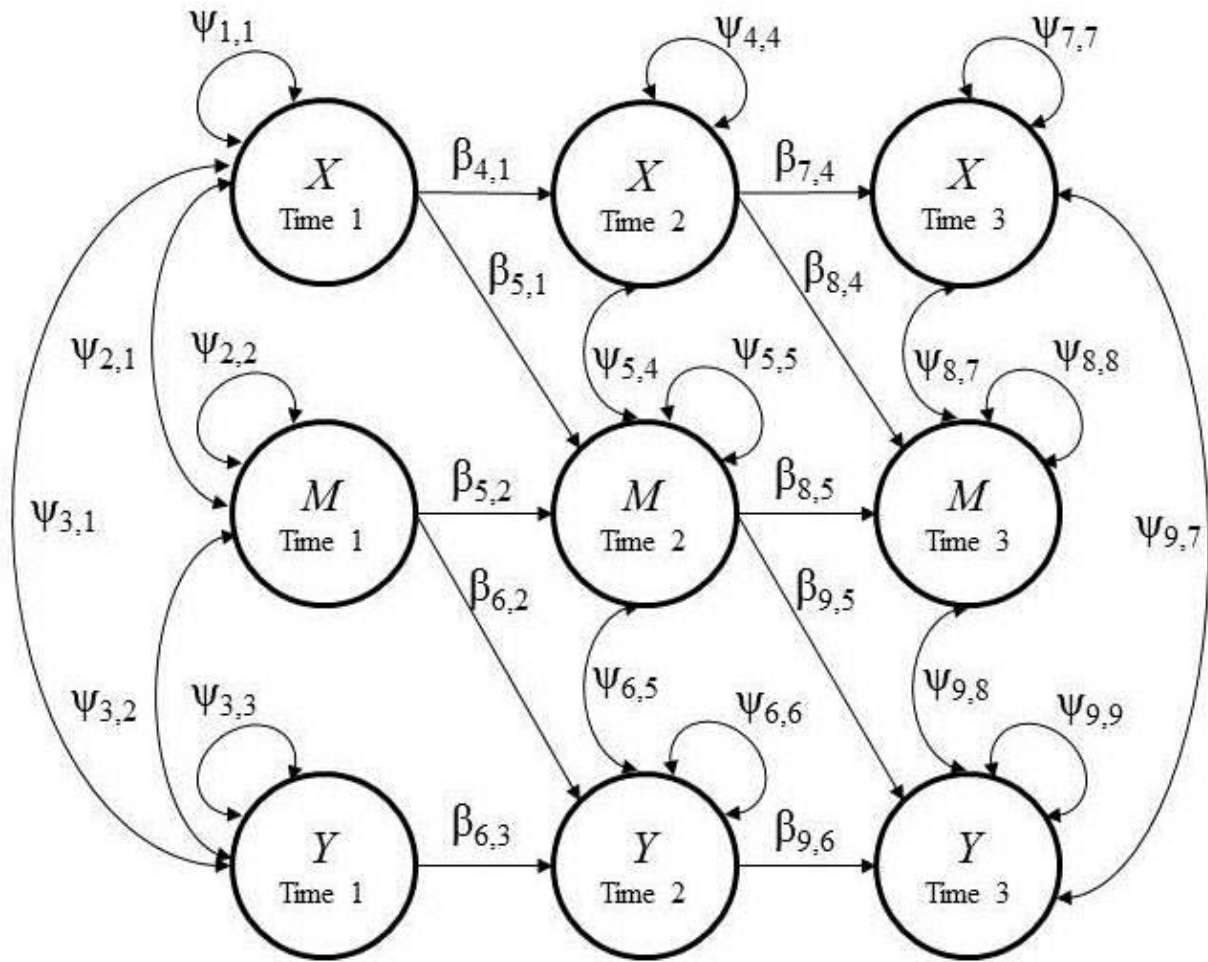


Figure 4: Data-generating and analysis model for Part II. To save space, only the structural model is depicted, and the within-time factor correlation $\psi_{6,4}$ is omitted at Time 2, although all are modeled. The measurement model (not depicted) includes three indicators per factor, with Lag-1 and Lag-2 residual correlations for the same indicators measured at each occasion.

I generated 200 data sets from each population. For each replication, I imposed 11.1% or 22.2% planned missing data using three different assignment methods (see patterns in Figures 1–3), and fit models to each of the six missing data sets. Planned missing data was simulated using a 2 (number of items in the X block) \times 3 (longitudinal assignment strategy) design. There were

either three items in the X block and two items in each of the A, B, and C blocks at each time point (“X3”; 22.2% missing data), or six items in the X block and one item in each of the A, B, and C blocks (“X6”, 11.1% missing data). Indicators of the same factor were never assigned to the same A, B, or C block because an ideal design would distribute as evenly as possible the missing information about each construct, rather than have the same subject miss more information about one construct than another (Raghunathan & Grizzle, 1995; Rhemtulla, Savalei, & Little, 2013). Assignment of items to blocks at any occasion is depicted in Figure 5 for each proportion of missing data condition (X3 or X6).

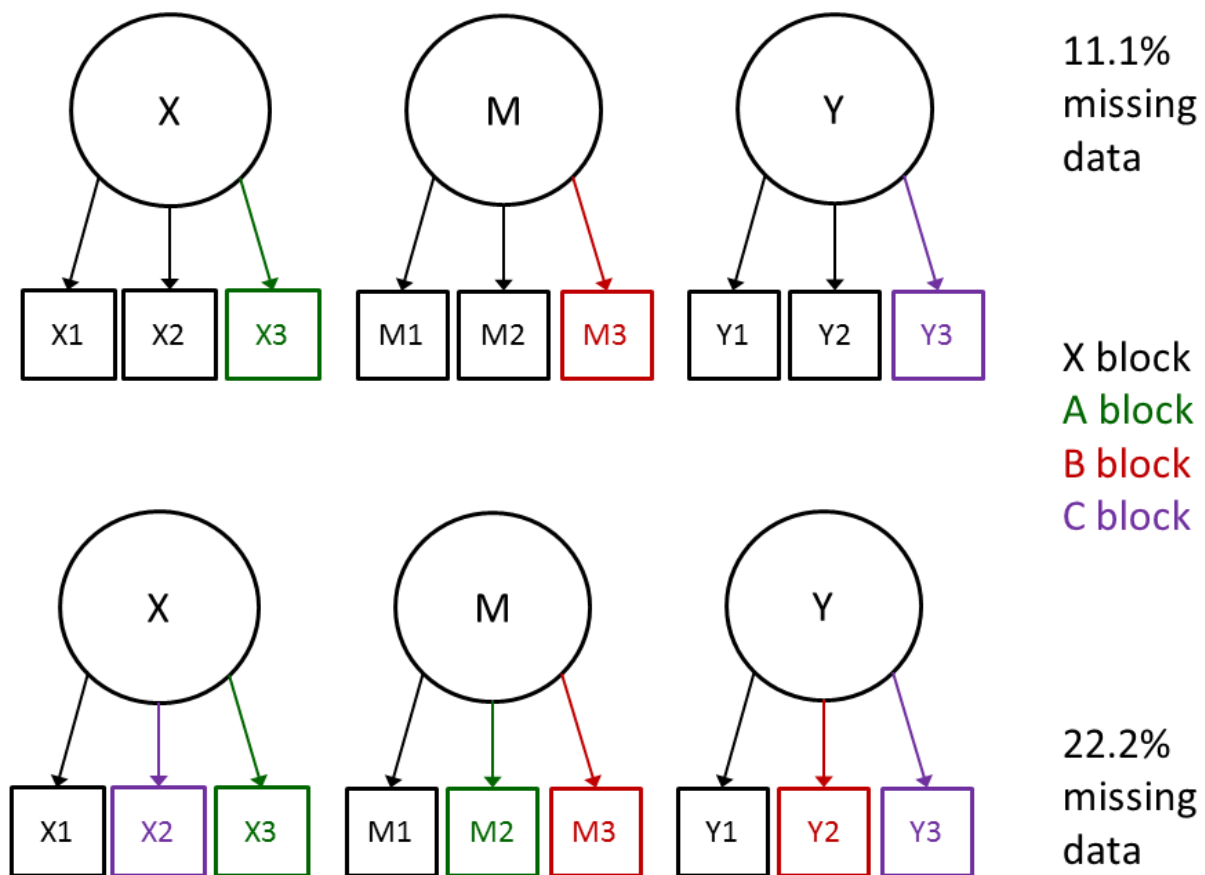


Figure 5: Items within a construct are evenly distributed across blocks within a given occasion.

Recall that assignment strategy refers to whether participants were assigned to complete the same form or different forms at each wave, or to be randomly assigned forms on each occasion. Because there were three forms to distribute over three occasions, participants in the different-forms condition completed each combination once: for example, (1) XAB, (2) XAC, (3) XBC. In the random condition, subjects had a one third chance of seeing any form on any occasion, regardless of what they had seen before, so there were 27 combinations of orders: for example, (1) XAB, (2) XBC, (3) XBC, or (1) XAC, (2) XBC, (3) XAC. On average, one ninth of subjects would see the same form on each occasion (like the same-form condition), two ninths would never see the same form twice (like the different-forms condition), two ninths would see the same form on nonconsecutive occasions (i.e., Times 1 and 3), and the remaining four ninths would see the same form on only one pair of consecutive occasions.

All data were simulated and fit using the R package *simsem* version 0.4-1 (Pornprasertmanit, Miller, & Schoemann, 2012), which is designed for simulating and analyzing data in an SEM framework. The *simsem* package has advanced missing data simulation capabilities, including the ability to simulate MCAR, MAR, and planned missing data patterns. Models in *simsem* are fit using *lavaan* (Rosseel, 2012), and missing data are handled either with FIML (a feature available in *lavaan*) or with multiple imputation using the R package *Amelia* (Honaker, King, & Blackwell, 2011). In my simulations in Parts II and III, all missing data were estimated using FIML. Six missing data conditions in each of 436 populations yields $436 \times 2 \times 3 = 2616$ conditions, 200 replications of which yield $200 \times 2616 = 523,200$ observed sets of results for analysis. I provide R syntax for one condition in Part II as a template for replication in Appendix A.

Results

In each planned-missing condition (X3 vs. X6, three assignment methods), approximately 97% of models converged on a proper solution. There were no differences in convergence rates or patterns of convergence problems between the different assignment methods. The remaining 3% of models converged on an improper solution that included a negative residual latent variance. As would be expected, improper solutions occurred in populations with residual latent variances close to zero (e.g., $\psi_{5,5}$ in Figure 1 $< .05$), but they occurred slightly less frequently when populations had higher factor loadings. Because the reason for negative residual variances is explained by population values close to zero, and no other problem was found with any solutions, I included all 523,200 replications in the results.

Parameter Estimates. I calculated five outcome measures for all 103 parameter estimates across 436 populations. Absolute bias in the parameter estimates is the difference between the average parameter estimate and the corresponding population value of that parameter. Because absolute bias is difficult to interpret due to the wide range of possible values, I also provide relative bias, which is the ratio of the absolute bias to the population value. Hoogland and Boomsma (1998) proposed that acceptable levels of relative bias have an absolute value < 0.05 . Likewise, relative *SE* bias is the ratio of (a) the difference between the average *SE* of the estimate across replications and the *SD* of the parameter estimate and (b) the *SD* of the parameter estimate. Hoogland and Boomsma proposed that acceptable levels of relative *SE* bias have an absolute value < 0.10 . The mean squared error ($MSE = \text{squared } SD \text{ of the estimate plus squared absolute bias of the estimate}$) is a combination of bias and variability of the estimates, the square root of which provides an estimate of the average distance from the population value that one might find in a given replication. This interpretation is similar to *RMSE* of the residuals

in a regression model: an estimate of how much an observed value can be expected to vary from the true population value. The proportion of replications for which the true population parameter is captured by the 95% CI is called the 95% coverage rate, which in the absence of bias is expected to be close to 95%. All of these estimates except for *MSE* are calculated automatically in *simsem*.

Table 2 provides the 5th and 95th empirical percentiles of bias, *MSE*, and coverage rates in each planned-missing condition. This gives an idea of how small the bias and *MSE* are, and of how large the coverage rates are, across the 103 parameter estimates and 436 populations. The columns for the different methods of assignment have almost identical ranges, and the mean difference (not shown) between any two methods is practically zero for all parameters. Upon closer inspection of individual populations (436 are too many to summarize individually here), almost all values of relative bias in estimates and *SE* were within the bounds suggested by Hoogland and Boomsma (1998). A few exceptions are discussed under Improper Solutions.

The lack of differences can be seen clearly by comparing top and bottom rows of Figures 6 and 7. Figure 6 shows the range of estimates for factor loading $\lambda_{14,5}$ as the population value increases. This loading was chosen for comparison across conditions because it is one of the items assigned to the X block (complete data) in the X6 condition, but was assigned to the A block (33% missing) in the X3 condition. The ranges in the bottom panels are not noticeably wider than those in the top panels, even though the estimates are made with fewer observations. No differences emerged for the regression of outcome *Y* at Time 3 on mediator *M* at Time 2 (parameter $\beta_{9,5}$ in Figure 4), which is partially defined by the factor loading in Figure 6.

Table 2

Range of Bias, MSE, and 95% Coverage in Six Planned-Missing Conditions

Number of Items in		Empirical 5 th and 95 th Percentiles		
the X Block	Measure	Different Forms	Same Forms	Random Forms
Six (11% Missing)	Absolute Bias	[−0.0076, 0.0058]	[−0.0015, 0.0011]	[−0.0020, 0.0037]
	Relative Bias	[−0.0131, 0.0480]	[−0.0097, 0.0036]	[−0.0120, 0.0279]
	Relative <i>SE</i> Bias	[−0.0085, 0.0145]	[−0.0088, 0.0140]	[−0.0116, 0.0126]
	<i>MSE</i>	[0.0006, 0.0041]	[0.0006, 0.0041]	[0.0006, 0.0042]
	95% Coverage	[0.9423, 0.9519]	[0.9445, 0.9527]	[0.9446, 0.9523]
Three (22% Missing)	Absolute Bias	[−0.0053, 0.0065]	[−0.0018, 0.0011]	[−0.0034, 0.0105]
	Relative Bias	[−0.0306, 0.0398]	[−0.0136, 0.0023]	[−0.0248, 0.0541]
	Relative <i>SE</i> Bias	[−0.0106, 0.0117]	[−0.0109, 0.0121]	[−0.0132, 0.0137]
	<i>MSE</i>	[0.0007, 0.0046]	[0.0007, 0.0046]	[0.0007, 0.0047]
	95% Coverage	[0.9427, 0.9517]	[0.9446, 0.9522]	[0.9433, 0.9518]

Note. Measures were calculated across 87,200 replications (200 within each of 436 populations). There were 88 parameter estimates in the covariance structure and 15 in the (just-identified) mean structure.

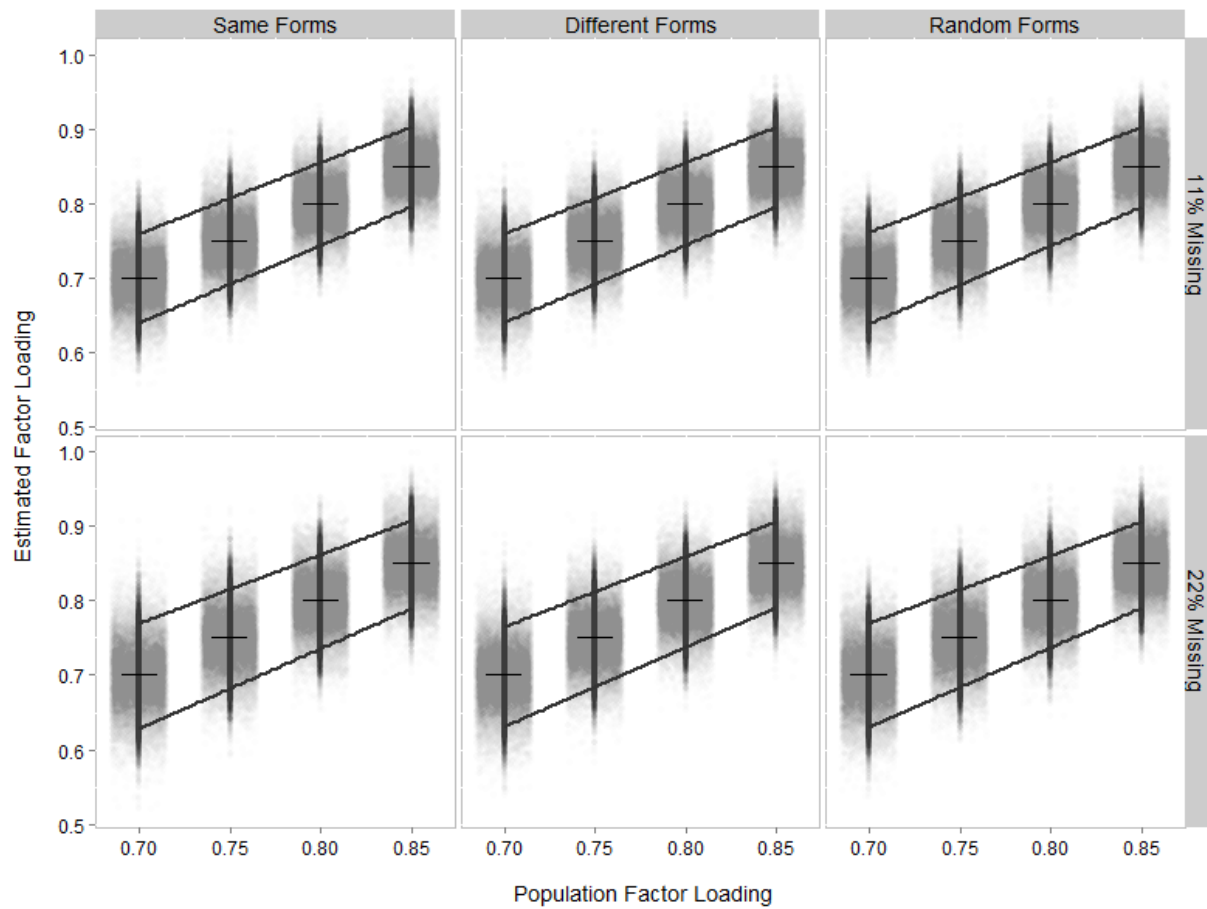


Figure 6: Variability in estimated factor loading across conditions. Parallel diagonal lines are empirical 90% confidence bands. The values were “jittered” along the x -axis in order to make it easier to see the density of observed estimates at each population value. I chose the second factor loading of the mediator at Time 2 because it had no missing values in the X6 condition, but one third missing in the X3 condition.

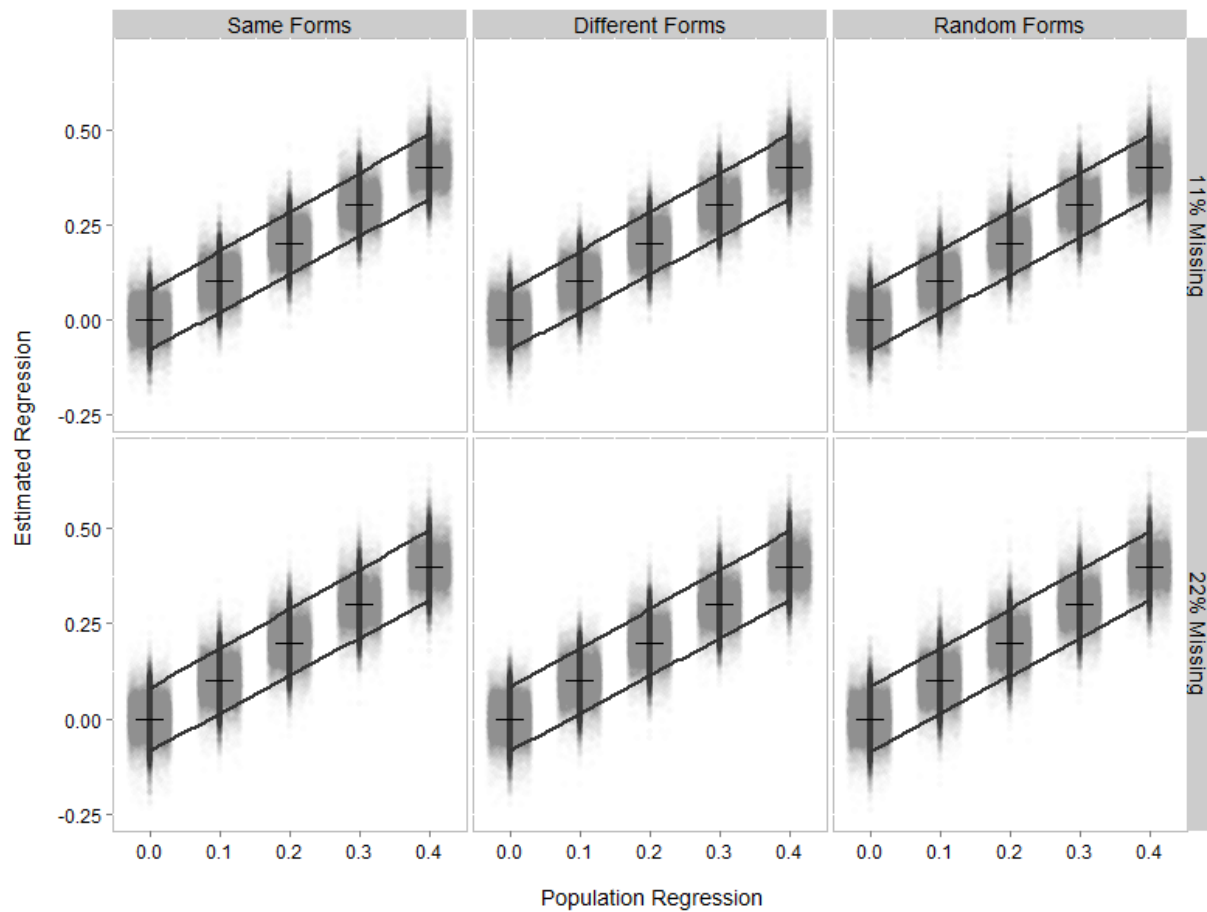


Figure 7: Variability in estimated factor regression across conditions. Parallel diagonal lines are empirical 90% confidence bands. The values were “jittered” along the x -axis in order to make it easier to see the density of observed estimates at each population value. I chose the regression of the outcome Y at Time 3 on the mediator M at Time 2, which is partially defined by the factor loading in Figure 2.

In a real analysis situation, the latent regression parameters would probably be of greater substantive interest. The population parameters for the latent regressions ranged from 0.0–0.4 in increments of 0.1, and Figure 8 depicts the increase in power with each increment. When the population parameter is zero, the Type I error rate is very close to nominal (i.e., 5% when $\alpha =$

.05). Power increases to 80% when the standardized population value is as little as 0.2, and is close to 100% for larger values, due at least in part to the large sample size ($N = 540$). Methods for determining sample size for desired power given a particular model are discussed in Muthén and Muthén (2002) and in Schoemann, Miller, Pornprasertmanit, and Wu (2013).

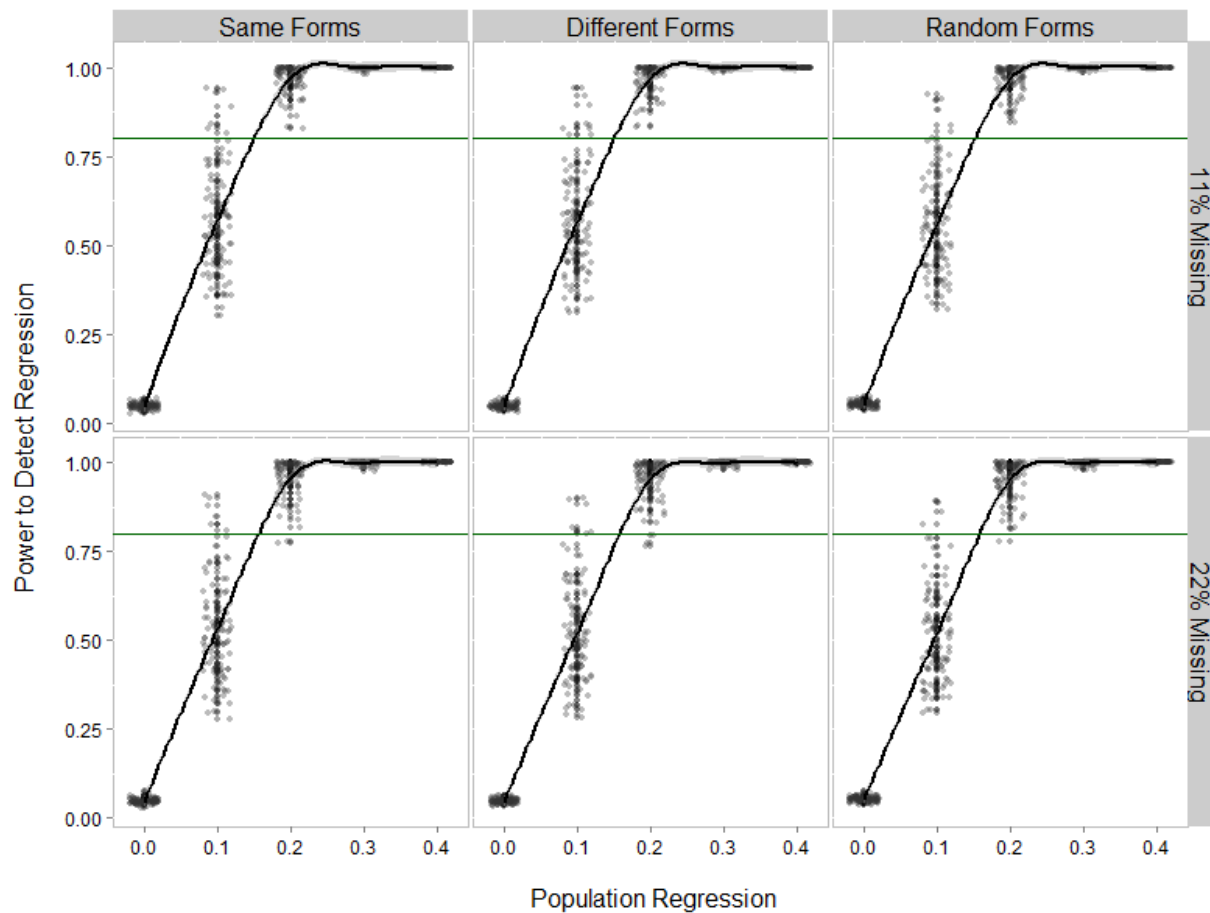


Figure 8: Power to detect latent regressions increases with effect size. The values were “jittered” along the x -axis in order to make it easier to see the density of observed power at each population value. When the population value is at least 0.2 (standardized), the power is greater than 80%. When the population value is zero, the Type I error rate is 5%.

Improper Solutions. I separately investigated the conditions when residual latent variances were below 0.05 in the population (on a standardized metric) because they led to many improper solutions. Still, no differences between assignment methods were detectable in these circumstances, although more extreme values were found in the tails of the distributions of relative bias in estimates (but not in *SEs*). Here, small differences between estimated and population coefficients can appear very large relative to very small population values, which was the case for the estimated latent residual variances and covariances in these populations. I do not see this as problematic, for three reasons. First, the large relative bias is purely an artifact of its being a ratio with a small value in the denominator. Second, neither the relative *SE* bias nor the 95% coverage rates are affected by relative bias (i.e., relative *SE* bias still falls below 0.10 for such parameter estimates). Third, longitudinal latent-variable models include estimates of residual covariances because there is reason to suspect that item-specific variance present at any occasion will also be present at all occasions (e.g., negatively worded items will remain negatively worded items). They are typically not included to be interpreted, so any degree of relative bias in such parameters is likely of little interest in an applied research situation. As long as the only sign of impropriety is a small, negative residual variance—only for a variable for which it would be expected to be close to zero—the solution is justifiably acceptable under such circumstances (Savalei & Kolenikov, 2008).

PART III: Differences among Assignment Methods in the Presence of Practice Effects

Modeling Practice Effects

Although the different assignment methods yield equivalent results in the absence of any retest effects, such a situation is not likely to occur in practice. There are many ways for retest effects to manifest, such as higher correlations among items repeatedly measured, especially

when repeated measurements are very proximal. Even if retest effects were confined to mean differences, they might have a complex pattern, such as greater change with more proximity, or a nonlinear increase that levels off after the second or third measurement (i.e., the retest effect might accumulate only up to a certain point).

Freund and Holling (2011) suggested three possible ways practice effects manifest: increases in factor (i.e., true score) values, a reduction in error confounding the measure, or increases to test-specific skills (i.e., systematic error, or item-specific uniqueness). Increases in factor scores would imply that practicing certain skills augments a person's ability. For example, if a participant presents a practice effect after multiple math tests, it is likely because their relevant mathematics skills have increased. A reduction of error confounding the measure implies that multiple measures reduces the measurement error present in data observations, leading to a less obscured, and subsequently greater, true score. Increases to test-specific skills does not imply that there is a change in a subject's latent traits or abilities, but that their scores increase as a result of improved performance ability on the specific test.

The best way to model practice effects often depends on what is measured. For example, complex cognitive models are not effectively explained with increasing factor values; they appear more related to increases of test-specific factors (i.e., verbal or quantitative reasoning, processing speed, etc.). Matton, Vautier, and Raufaste (2009) attributed this change to the situational component of scores that is believed to effectively encompass any variation in retesting situations. I chose this type of practice effect—incremental increases in the means of items that are seen on consecutive occasions, keeping the corresponding error variances and true-score latent means constant—for a simple, straightforward proof of concept that assigning different forms over time is likely to be preferred in the presence of any type of retest effect.

Monte Carlo Simulation Design

I simulated practice effects in the form of an incremental increase in the mean (indicator intercept) by 0.1 for any variable that was seen on consecutive occasions. Because these were standard normal variables, this corresponds to a Cohen's $d = 0.1$, which is a small effect (Cohen, 1988). An increasing mean on a measure mimics an ability measurement, such that more frequent exposure yields greater improvement due to practice. All variables in Block X were seen on each occasion, so the means for those indicators were 0.0, 0.1, and 0.2, which is the same pattern for all items in Blocks A–C when the same form was assigned each time. Assigning a different form each time minimizes this effect so that, for example, only items in Blocks X and A would show evidence of reactivity if a subject was assigned to form XAB followed by XAC. Assigning random forms as described above would minimize reactivity for (on average) four ninths of subjects to the same degree, but would do so less effectively for another four ninths and not at all for one ninth.

To demonstrate how assigning different forms over time can attenuate practice effects, I fit a three-factor CFA that represented a single factor measured across three waves. I chose a CFA instead of the mediation model in Part II because without latent regression parameters, I could directly estimate latent means rather than calculating them from latent intercepts and regressions. Each latent variable had seven indicators at each wave, allowing me to increase the proportion of missing data from Part II (see Figure 9). Similar to Part II, the population measurement model included Lag-1 residual correlations of 0.2 and Lag-2 residual correlations of 0.04 across time, and the analysis model estimated residual covariances at Lags 1 and 2. Strong factorial invariance was again established by constraining factor loadings and intercepts to equality across time in the analysis model, and the analysis model was identified by fixing the

latent variance and mean at Time 1 to 1 and 0, respectively. Latent means remained zero in the population, so any systematic increase in estimates of latent means would indicate bias due to practice effects. I also fit a weak-invariance model to the same data sets, with which I verified that the intercepts themselves were biased when all latent means were fixed to zero. Factor loadings varied from 0.7 to 0.85, by increments of 0.05, and factor correlations varied from 0.4 to 0.9 by increments of 0.1. This resulted in $4 \times 6 = 24$ population models. Sample size for all replications in Part III was fixed to $N = 270$.

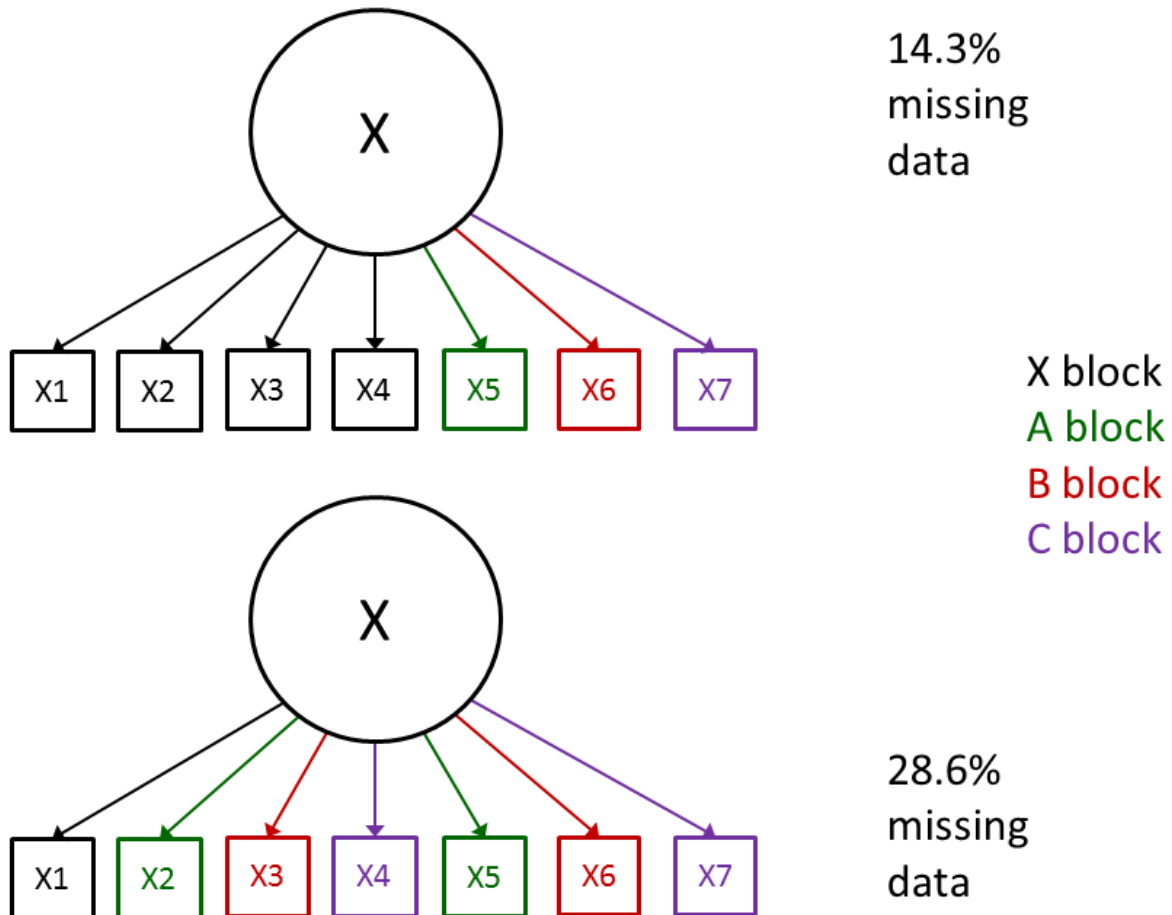


Figure 9: Assignment of items to blocks at one occasion.

For each replication, I imposed planned missing data using the same assignment strategies from Part II, and I fit models to each of six missing data sets. Because there were seven indicators, the first indicator was assigned to the X block in all conditions. In the 14.3% missing conditions (X4), indicators 2–4 were also assigned to the X block and indicators 5–7 were assigned to Blocks A, B, and C, respectively (see Figure 9). In the 28.6% missing conditions (X1), indicators 2–4 were assigned to Blocks A, B, and C, respectively. All data were simulated and analyzed using *simsem* and *lavaan*, respectively, and missing data were estimated using FIML. Six missing data conditions in 24 populations yields $24 \times 2 \times 3 = 144$ conditions, 200 replications of which yield $200 \times 144 = 28,800$ observed sets of results for analysis. I provide R syntax for one condition in Part III as a template for replication in Appendix B.

Results

Because I analyzed a CFA model, all latent-variable variances were exogenous, and there were no population residual variances close zero. Consequently, 100% of analyzed models converged on proper solutions. Similar to Part II, analysis of the parameter estimates in the covariance structure shows that assignment methods yield essentially equivalent results, given the absence of any effects (on the covariance structure) of repeated sampling. Thus detailed tables and figures would be redundant.

Analysis of the mean structure, on the other hand, shows how assigning different forms over time would be preferable when practice effects are present. Intercepts from the weak-invariance model are presented in Figure 10, along with the latent means from the strong-invariance model, wherein the latent means were freely estimated at Times 2 and 3 (see Little, 2013). The intercepts have identical behavior when assigning the same forms over time because participants would see the same items on each occasion, regardless of whether the item was in

the X block or the ABC blocks. When the intercepts are constrained to equality, this bias is manifested in the latent means, depicted with the bold black line. Random assignment of forms over time results in less bias, especially with more variables in the ABC blocks (i.e., more planned missing data, as shown in the X1 condition compared to X4), but is not as noticeable of an improvement as systematically assigning different forms each time.

PART IV: Conclusions

Suggestions for Applied Researchers

In the absence of retest effects, there is little difference between methods, which suggests that any extra expense of assigning different forms (randomly or systematically) would have no benefit. However, even in the unlikely situation where retest effects are absent, assigning different forms over time has no apparent statistical drawbacks and can be substantively justified when, for example, measuring multidimensional constructs or using scales whose items are not tau-equivalent (items with equal reliability, i.e., essentially equal factor loadings, are tau-equivalent). For example, an intelligence measure might include an item about verbal, quantitative, and spatial intelligence, each of which measures a different aspect of a general intelligence construct. In such a situation, I recommend gathering as much information about a person's standing on a construct as possible by assigning different items on each occasion. Assigning the same form over time would result in measuring only verbal intelligence for one third of participants, only quantitative for another third, etc. Of course, each dimension of a general intelligence construct would typically have multiple indicators that can be evenly dispersed among planned-missing forms, just as items of more unidimensional constructs should be evenly dispersed among planned-missing forms (Graham et al., 1996, 2006).

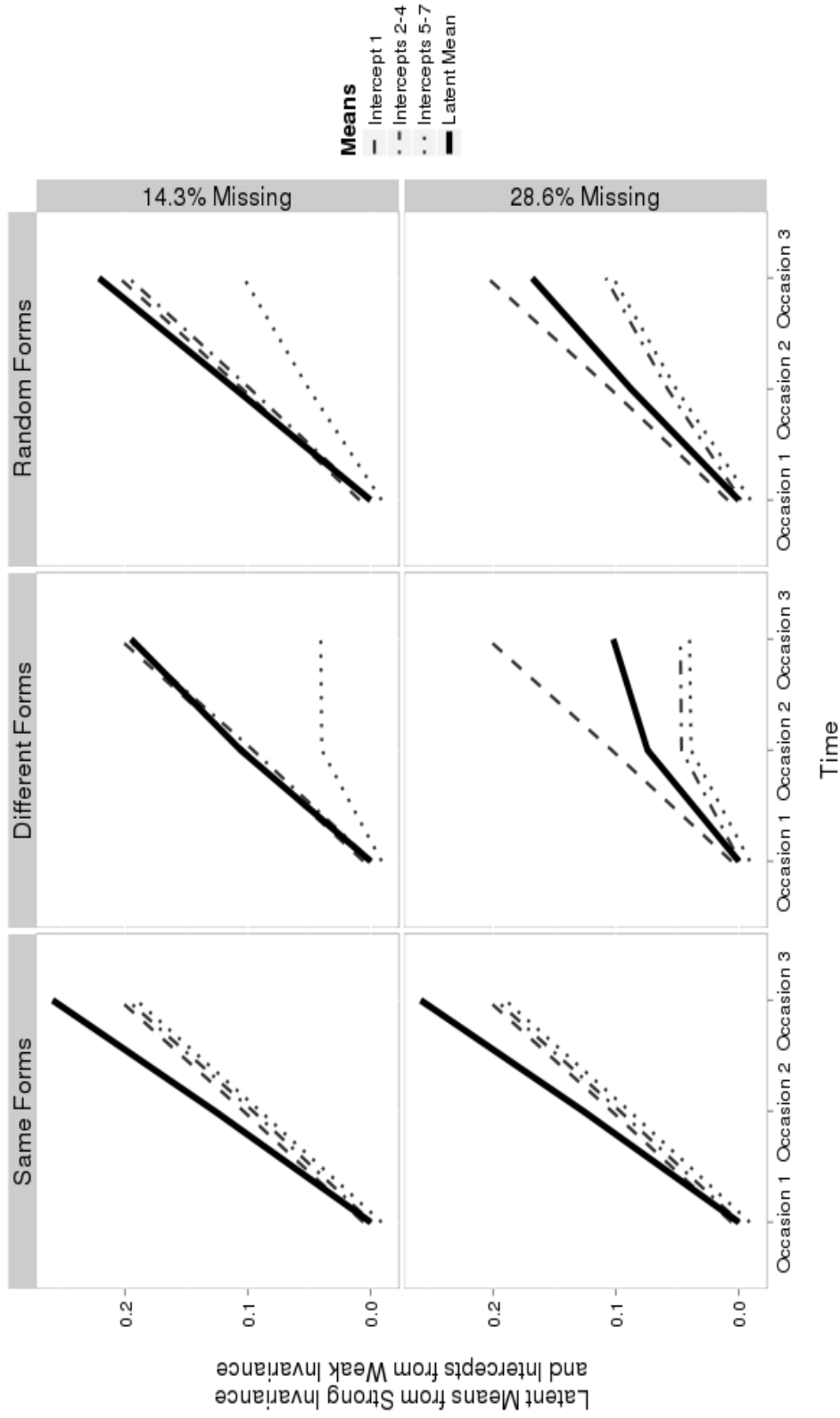


Figure 10: Bias in latent means and intercepts due to practice effects over time. The true latent mean is zero, and the intercepts begin at zero and increase by 0.1 when seen on consecutive occasions. The intercept values were taken from a weak-invariance model, in which all latent means were fixed at zero and all intercepts were freely estimated. The latent means were taken from the strong-invariance model, in which intercepts were constrained to equality over time and latent means were freely estimated after Time 1.

The more likely situation is one in which retest effects of some kind are present, and my research has shown how measurement-related bias can also manifest in latent parameter estimates. I simulated only one kind of practice effect—increasing means for indicators seen on consecutive occasions—but practice effects can also manifest in the covariance structure (e.g., stronger correlations among items seen on consecutive occasions) and have more complicated patterns. Regardless of the form of retest effect, assigning different forms over time can alleviate bias to a degree not possible when assigning the same form over time.

Rather than try to find a generalizable rule of thumb using only one kind of model and one way of simulating a particular kind of practice effect, I simply point out that even when the true latent trait does not change over time, it is probable that practice effects will contaminate the latent parameter estimates. The goal is to minimize this contamination. Any longitudinal design that involves repeated measures could suffer from this source of bias, but the three-form planned missing design can minimize this bias if it is implemented properly. Assigning the same forms over time will provide the general benefits of lowering costs, response burden, and fatigue, but minimizing bias due to retest effects requires different forms to be assigned over time. Random assignment of forms will reduce bias to some degree, but it would likely be worth the additional effort to ensure that participants cycle through each form before seeing their first one again.

Systematically assigning different forms over time would always result in less bias due to retest effects, but the degree to which it is more effective than random assignment may not necessarily outweigh the additional cost of the systematic approach in certain situations. For instance, if the survey is administered online on each occasion, it may be a very simple matter to program the web application to present all participants with the X block first, then randomly present additional items from the A, B, or C blocks. Indeed, such randomization is a feature in

some web sites for distributing surveys online. In such situations, random assignment may be the most efficient use of time, and it still protects against bias in a way that assigning the same form does not. In many other situations, however, it may actually take more effort for researchers to devise a method of random assignment than it would to simply assign each group of participants to a particular fixed pattern at the beginning of the study, one which would assure they would not see the same form until they had seen each other form. Systematic assignment would then both be more time-efficient and provide better protection against bias.

It is worth reiterating some general advice about multiform designs provided by Graham et al. (2006). Researchers should include in the X block demographic variables, any dependent variables not measured with multiple indicators, and any variables that are expected to be related to unplanned missingness (e.g., SES could be a predictor of attrition). My simulations placed at least one item from each construct in the X block, but this is not necessary if there are many indicators with high loadings, because numerous items would be capable of measuring the construct equivalently well, making them interchangeable for practical purposes. In the case of a scale with low reliability, it may be wise to put in the X block the item with the highest reliability or that is most representative of the construct. Regardless, items within a construct should be divided among the different item sets.

Greater proportions of missing data can be achieved with other multiform designs, which might result in greater cost efficiency, depending on the relative costs of recruiting participants versus the time spent acquiring measurements. For instance, a 10-form design has an X block and A–E blocks, and participants are assigned items from X and any of 10 possible pairs of sets from A–E (e.g., XAD, XBE, etc.). Studies with many measurements may benefit from this more involved multiform design, but researchers should take into account their planned (and

unplanned) missingness when conducting power analyses to ensure a sufficient sample size to detect effects of interest. My studies only involved two kinds of models, but the *simsem* software (Pornprasertmanit et al., 2012; Schoemann et al., 2013) can be used to conduct power analyses for numerous types of structural models (*simsem* is a free open-source R package).

Limitations and Future Directions

In Parts II and III, the analysis models matched the population models, so I had no conditions with any kind of misspecification. Thus I am not able to draw conclusions about the effect of omitted parameters or omitted variables on the ability of a three-form planned missing design to allow for unbiased estimates. It is unclear whether such effects on bias (or model fit, etc.) in the context of planned missing designs would differ from a complete-data context, but in general, one can assume that adding MCAR missingness will reduce parameter efficiency (power) by increasing *SEs*, but not affect bias. There is no reason to suspect that this feature of MCAR missingness would be any different when there is some kind of model misspecification.

Although I demonstrate that in the absence of retest effects all assignment methods yield similar results, my constructs were unidimensional in the population. Multidimensional constructs may make it preferable to assign different forms over time even in the absence of retest effects so that participants can respond to a greater range of scale items. Certain key scale items from such constructs may be placed in the X block so that researchers are certain to measure that full range. Because a complete lack of retest effects of any kind is unlikely in any situation, assigning different forms is preferred even for unidimensional constructs. Still, the case of multidimensional constructs may be worth investigating in future research.

Related to the omitted variable problem is the MNAR mechanism. I simulated only planned missingness, which is an MCAR mechanism when participants are randomly assigned to

different forms (on the first occasion, regardless of which of the discussed assignment strategies is used for later occasions). Although planned missing designs have the potential to reduce unplanned missingness, some proportion of observations missing due to MCAR, MAR, or MNAR mechanisms is likely. As discussed above, MNAR data can become MAR data with conscientious planning, by including variables that are related to attrition or nonresponse.

I simulated latent-variable models, which have multiple indicators per construct. Thus, my results may not generalize to different types of models that use single indicators, such as some latent growth curve models, path analyses, multiple regressions, or multilevel regressions. My factors were unidimensional, and my indicators were tau-equivalent, so it essentially did not matter which variables were assigned to the X or ABC blocks. Tau-equivalence is a rarely met assumption in practice, which could have undesired consequences on the outcome of an analysis, so it is an important topic for further research. Likewise, it is common for multiple factors to account for an indicator's variance when there are expected cross-loadings or method effects (e.g., multitrait–multimethod models). Further investigation into the nuances of such study designs is warranted; however, it is difficult to address a wide range of possibilities in a single study. I therefore reinforce my suggestion that researchers would be wise to take the nuances of their own designs into account by incorporating these nuances (along with planned missingness) into a simulation-based power analysis, which could also shed light on potential convergence or identification issues that might not be immediately apparent.

I analyzed all models using FIML rather than MI. When the imputation and analysis models are equivalent, pooled results from larger numbers of imputations converge to identical results using FIML. Because my missing values are planned (i.e., MCAR), MI is expected to yield equivalent results. I kept my moderate-to-large sample sizes fixed, but other problems may

arise with smaller samples that I did not discover. For instance, all models converged, but not all of them on proper solutions. Although the negative residual variances only occurred when the population value was close to zero, other reasons for nonconvergence might arise with smaller samples. Additional (unplanned) missing data increases the number of missing data patterns, and might become particularly problematic when forms are randomly assigned over time because that also increases the number of missing data patterns. In such situations, MI might provide a robust alternative to FIML because analyses would be conducted on imputed samples (i.e., complete data) before final results are combined across imputations.

In conclusion, latent-variable models appear robust to assignment methods only in the absence of retest effects and when indicators within a construct are distributed evenly across planned-missing forms. Because retest effects should realistically be anticipated in many situations, I recommend systematically assigning different forms over time, rather than randomly assigning forms, to ensure that measurements on the same items are repeated as seldom as possible, especially on proximate occasions.

References

- Anderson, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association*, 52(278), 200–203. doi:10.1080/01621459.1957.10501379
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48, 5–37. doi:10.1016/j.jsp.2009.10.001
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, 112, 558–577. doi:10.1037/0021-843X.112.4.558
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Freund, P. A., & Holling, H. (2011). How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items. *Intelligence*, 39(4), 233–243. doi:10.1016/j.intell.2011.02.009
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.
doi:10.1007/978-1-46144018-5_1
- Graham, J. W., Hofer, S. M., & MacKinnon, D. P. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197–218.
doi:10.1207/s15327906mbr3102_3
- Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, 11(4), 323–343.
doi:10.1037/1082-989X.11.4.323

- Harel, O., Stratton, J., & Aseltine, R. (2011). *Designed missingness to better estimate efficacy of behavioral studies* (Technical Report 11-15). Storrs, CT: Department of Statistics, University of Connecticut.
- Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45, 1–47. Available from <http://www.jstatsoft.org/v45/i07>
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. *Sociological Methods & Research*, 26, 329–367. doi:10.1177/0049124198026003003
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford.
- Matton, N., Vautier, S., & Raufaste, E. (2009). Situational effects may account for gain scores in cognitive ability testing: A longitudinal SEM approach. *Intelligence*, 37(4), 412–421. doi:10.1016/j.intell.2009.03.011
- Pornprasertmanit, S., Miller, P., & Schoemann, A. M. (2012). simsem: SIMulated structural equation modeling (version 0.4-1) [R package]. Available from the Comprehensive R Archive Network: <http://cran.r-project.org/>
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, 9, 599–620. doi:10.1207/S15328007SEM0904_8
- Raghunathan, T. E., & Grizzle, J. E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association*, 90, 54-63. doi:10.1080/01621459.1995.10476488
- Rhemtulla, M., Savalei, V., & Little, T. (2013). *On the asymptotic relative efficiency of planned missingness designs*. Manuscript in preparation.

- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1–36. Available from <http://lavaan.ugent.be/>
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Hoboken, NJ: Wiley.
- Savalei, V., & Kolenikov, S. (2008). Constrained versus unconstrained estimation in structural equation modeling. *Psychological Methods*, 13, 150–170. doi:10.1037/1082-989X.13.2.150
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7, 147–177. doi:10.1037//1082-989X.7.2.147
- Schoemann, A. M., Miller, P. R., Pornprasertmanit, S., & Wu, W. (2013). *Using Monte Carlo simulations to determine power and sample size in planned missing data designs*. Manuscript submitted for publication.
- van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.

Appendix A

R Syntax for One Condition in Part II: Equivalence of Assignment Methods

```

library(simsem)

## population parameters
fl <- 0.7
wtc <- 0.2
ar1 <- 0.5
cl1 <- 0.3
## choose residual variances that set total indicator variance equal to 1
resvar <- 1 - fl^2
## choose latent residual variances that set total factor variance equal to 1
t2xvar <- 1 - (ar1^2)
t2mvar <- 1 - (cl1^2 + ar1^2 + 2*(ar1 * wtc * cl1))
t2yvar <- 1 - (cl1^2 + ar1^2 + 2*(ar1 * wtc * cl1))

## LAMBDA matrix
makeLambda <- function(inpMat, nFac = 0, nTimes = 0, npf = 0, val) {
  val <- rep(val, length.out = prod(nFac, nTimes, npf))
  for (i in 1:(nFac * nTimes)) {
    a <- ((npf * (i - 1)) + 1)
    b <- (((npf * (i - 1)) + npf))
    inpMat[a:b, i] <- val[a:b]
  }
  return(inpMat)
}

loading <- matrix(0, 27, 9)
loading <- makeLambda(loading, 3, 3, 3, paste0("LY", 1:9)) # weak invariance

load.val <- matrix(0, 27, 9)
load.val <- makeLambda(load.val, 3, 3, 3, fl)

LY <- bind(loading, load.val)

## THETA matrix
errorLag <- function(inpMat, nVar = 0, nTime = 0, nlag = 0, val) {
  i <- 1
  while (i <= ((nVar * nTime) - (nlag * nVar))) {
    inpMat[i, (i + (nlag * nVar))] <- val
    inpMat[(i + (nlag * nVar)), i] <- val
    i <- i + 1
  }
  return(inpMat)
}

error.na <- matrix(0, 27, 27)
diag(error.na) <- NA
error.na <- errorLag(error.na, 9, 3, 1, NA)
error.na <- errorLag(error.na, 9, 3, 2, NA)

error.cor <- matrix(0, 27, 27)
diag(error.cor) <- 1

```



```

error.cor <- errorLag(error.cor, 9, 3, 1, .2)

## this specifies residual covariance matrix as a correlation matrix
RTE <- binds(error.na, error.cor)
## this specifies the residual variances to rescale the above correlations
error.var <- rep(NA, 27)
VTE <- bind(error.var, rep(resvar, 27))

## PSI matrix
makePsi <- function(inpMat, nFac = 0, nTimes = 0, val) {
  if (length(val) == 1) {
    val <- rep(val, nFac)
  }
  if (length(val) == nFac) {
    for (i in 1:nTimes) {
      inpMat[(1 + (nFac * (i - 1))), (2 + (nFac * (i - 1)))] <- val[1]
      inpMat[(2 + (nFac * (i - 1))), (1 + (nFac * (i - 1)))] <- val[1]
      inpMat[(1 + (nFac * (i - 1))), (3 + (nFac * (i - 1)))] <- val[2]
      inpMat[(3 + (nFac * (i - 1))), (1 + (nFac * (i - 1)))] <- val[2]
      inpMat[(3 + (nFac * (i - 1))), (2 + (nFac * (i - 1)))] <- val[3]
      inpMat[(2 + (nFac * (i - 1))), (3 + (nFac * (i - 1)))] <- val[3]
    }
  }
  if (length(val) != nFac) {
    paste("Cannot evaluate: unequal val= and nFac=")
  } else {
    return(inpMat)
  }
}

factor.na <- matrix(0, 9, 9)
diag(factor.na) <- 1
factor.na <- makePsi(factor.na, 3, 3, NA)

factor.cor <- matrix(0, 9, 9)
diag(factor.cor) <- 1
factor.cor <- makePsi(factor.cor, 3, 3, wtc)

RPS <- binds(factor.na, factor.cor) # specifies residual correlations
## specify residual variances to rescale above matrix as factor covariances
facVar <- rep(NA, 9)
popParam <- c(rep(1, 3), rep(c(t2xvar, t2mvar, t2yvar), 2))
VPS <- bind(facVar, popParam)
VPS@free[1:3] <- 1

## BETA matrix
makeBeta <- function(inpMat, nFac = 0, nTimes = 0, val= c(x, m, y, a, b, cp))
{
  if (length(val) == 1) {
    val <- rep(val, 2*nFac)
  }
  if (length(val) == 2*nFac) {
    for (i in 2:nTimes) {
      inpMat[(1 + (nFac * (i - 1))), (1 + (nFac * (i - 2)))] <- val[1]
      inpMat[(2 + (nFac * (i - 1))), (2 + (nFac * (i - 2)))] <- val[2]
      inpMat[(3 + (nFac * (i - 1))), (3 + (nFac * (i - 2)))] <- val[3]
    }
  }
}

```

```

        inpMat[(2 + (nFac * (i - 1))), (1 + (nFac * (i - 2)))] <- val[4]
        inpMat[(3 + (nFac * (i - 1))), (2 + (nFac * (i - 2)))] <- val[5]
        inpMat[(3 + (nFac * (i - 1))), (1 + (nFac * (i - 2)))] <- val[6]
    }
}
if (length(val) != 2*nFac) {
  paste("Cannot evaluate: unequal val= and nFac=")
} else {
  return(inpMat)
}
}

path.na <- matrix(0, 9, 9)
path.na <- makeBeta(path.na, 3, 3, c(rep(NA, 5), 0))

path.st <- matrix(0, 9, 9)
path.st <- makeBeta(path.st, 3, 3, c(ar1, ar1, ar1, cl1, cl1, 0))

BE <- bind(path.na, path.st)

## Bind matrices together to specify full population/analysis model
mod <- model.sem(BE = BE, LY = LY, RPS = RPS, RTE = RTE, VPS = VPS, VTE= VTE)

## matrix of missing-data patterns
missMat <- matrix(FALSE, 540, 28)
missMat[ 1:180, c(AB1, AC2, BC3)] <- TRUE
missMat[181:360, c(AC1, BC2, AB3)] <- TRUE
missMat[361:540, c(BC1, AB2, AC3)] <- TRUE
missPattern <- miss(logical = missMat)

results <- sim(nRep = 200, model = analMod, generate = popMod, n = 180,
              miss = missPattern, datafun = oneGroup, seed = 3141593)

summary(results)
summaryParam(results)

```

Appendix B

R Syntax for One Condition in Part III: Differences among Assignment Methods

```

library(simsem)

## missing item(s) from each planned missing forms
AB1 <- 7
AC1 <- 6
BC1 <- 5
AB2 <- AB1 + 7
AC2 <- AC1 + 7
BC2 <- BC1 + 7
AB3 <- AB1 + 14
AC3 <- AC1 + 14
BC3 <- BC1 + 14

## parameters
fl <- 0.7
fc <- 0.5
## Residual variances set them equal to 1
resvar <- 1 - fl^2

## LAMBDA matrix
makeLambda <- function(inpMat, nFac = 0, nTimes = 0, npf = 0, val) {
  val <- rep(val, length.out = prod(nFac, nTimes, npf))
  for (i in 1:(nFac * nTimes)) {
    a <- ((npf * (i - 1)) + 1)
    b <- (((npf * (i - 1)) + npf))
    inpMat[a:b, i] <- val[a:b]
  }
  return(inpMat)
}

loading <- matrix(0, 21, 3)
loading <- makeLambda(loading, 1, 3, 7, paste0("LY", 1:7)) # weak invariance

load.val <- matrix(0, 21, 3)
load.val <- makeLambda(load.val, 1, 3, 7, fl)

LY <- bind(loading, load.val)

## THETA matrix
errorLag <- function(inpMat, nVar = 0, nTime = 0, nlag = 0, val) {
  i <- 1
  while (i <= ((nVar * nTime) - (nlag * nVar))) {
    inpMat[i, (i + (nlag * nVar))] <- val
    inpMat[(i + (nlag * nVar)), i] <- val
    i <- i + 1
  }
  return(inpMat)
}

error.na <- matrix(0, 21, 21)
diag(error.na) <- NA

```

```

error.na <- errorLag(error.na, 7, 3, 1, NA)
error.na <- errorLag(error.na, 7, 3, 2, NA)

error.cor <- matrix(0, 21, 21)
diag(error.cor) <- 1
error.cor <- errorLag(error.cor, 7, 3, 1, .2)

RTE <- binds(error.na, error.cor)

error.var <- rep(NA, 21)
VTE <- bind(error.var, rep(resvar, 21))

## PSI matrix
factor.na <- matrix(NA, 3, 3)
factor.na[1, 1] <- 1

factor.cor <- matrix(fc, 3, 3)
diag(factor.cor) <- 1

PS <- binds(factor.na, factor.cor)

## TAU vectors: 1 "population" for each assignment group in order to
## model practice effects (each group will have increasing intercept
## for different combination of variables)
diffAB <- rep(0, 21)
diffAB[setdiff( 8:14,c(AB2,AC2))] <- diffAB[setdiff( 8:14, c(AB2, AC2))] + .1
diffAB[setdiff(15:21,c(AB3,AC3))] <- diffAB[setdiff(15:21, c(AB3, AC3))] + .1
diffAB[setdiff(15:21,c(AC3,BC3))] <- diffAB[setdiff(15:21, c(AC3, BC3))] + .1

diffAC <- rep(0, 21)
diffAC[setdiff( 8:14,c(AC2,BC2))] <- diffAC[setdiff( 8:14, c(AC2, BC2))] + .1
diffAC[setdiff(15:21,c(AC3,BC3))] <- diffAC[setdiff(15:21, c(AC3, BC3))] + .1
diffAC[setdiff(15:21,c(BC3,AB3))] <- diffAC[setdiff(15:21, c(BC3, AB3))] + .1

diffBC <- rep(0, 21)
diffBC[setdiff( 8:14,c(BC2,AB2))] <- diffBC[setdiff( 8:14, c(BC2, AB2))] + .1
diffBC[setdiff(15:21,c(BC3,AB3))] <- diffBC[setdiff(15:21, c(BC3, AB3))] + .1
diffBC[setdiff(15:21,c(AB3,AC3))] <- diffBC[setdiff(15:21, c(AB3, AC3))] + .1

## population model (list of 3 models, one for each assignment method)
TYdiffAB <- bind(free = rep(NA, 21), popParam = diffAB)
TYdiffAC <- bind(free = rep(NA, 21), popParam = diffAC)
TYdiffBC <- bind(free = rep(NA, 21), popParam = diffBC)
TYdiff <- list(TYdiffAB, TYdiffAC, TYdiffBC)

## analysis models
TYweak <- bind(rep(NA, 21), rep(0, 21)) # weak invariance
TYstrong <- bind(rep(paste0("T", 1:7), 3), rep(0, 21)) # strong invariance
AL <- bind(c(0, NA, NA), c(0, 0, 0)) # latent means = 0, free after Time 1

## Bind matrices together to specify model
popMod <- model.cfa(LY= LY, PS= PS, RTE= RTE, VTE= VTE, TY= TYdiff, AL= AL)
weakMod <- model.cfa(LY = LY, PS = PS, RTE = RTE, VTE = VTE, TY = TYweak)
strongMod <- model.cfa(LY= LY, PS=PS, RTE=RTE, VTE=VTE, TY= TYstrong, AL= AL)

## matrix of missing-data patterns

```

```

missMat <- matrix(FALSE, 270, 22)
missMat[ 1:90 , c(AB1, AB2, AB3)] <- TRUE
missMat[ 91:180, c(AC1, AC2, AC3)] <- TRUE
missMat[181:270, c(BC1, BC2, BC3)] <- TRUE
missPattern <- miss(logical = missMat)

## change "group" variable in generated data sets so it is all 1 group
oneGroup <- function(dat) {
  dat$group <- 1
  dat
}

## fit weak model, save CFI/Chi-sq values and intercepts
results <- sim(nRep = 200, model = weakMod, generate = popMod, n = 90,
              miss = missPattern, datafun = oneGroup, seed = 1234567)
weakChi <- results@fit$Chi
weakCFI <- results@fit$CFI
weakDF <- results@fit$df
taus <- colMeans(results@coef[c(69, 70, 73, 76, 77, 80, 83, 84, 87)])

## fit strong model and test strong invariance
results <- sim(nRep = 200, model = strongMod, generate = popMod, n = 90,
              miss = missPattern, datafun = oneGroup, seed = 1234567)
CHI <- (results@fit$Chi - weakChi) < qchisq(.95, results@fit$df - weakDF)
CFI <- (weakCFI - results@fit$CFI) < .01

summary(results)
summaryParam(results)

mean(CFI) # how often does strong invariance fail using delt-CFI

```